DOCUMENT RESUME

ED 265 208 TM 860 075

AUTHOR	Stedman, Lawrence C.; Kaestle, Carl F.
TITLE	The Test Score Decline Is Over: A Reinterpretation.
INSTITUTION	Wisconsin Center for Education Research, Madison.
SPONS AGENCY	National Inst. of Education (ED), Washington, DC.
REPORT NO	WCER-PR-85-8
PUB DATE	Oct 85
GRANT	NIE-G-84-0008
NOTE	35p.
PUB TYPE	Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Comparative Testing; *Educational Assessment; *Educational Trends; Elementary Secondary Education; *National Norms; National Surveys; Outcomes of Education; Scores; Standardized Tests; *Test Interpretation; *Test Use
IDENTIFIERS	ACT Assessment; Iowa Tests of Basic Skills; National Assessment of Educational Progress; Scholastic Aptitude Test; Stanford Achievement Tests; *Test Score Decline; Tests of Achievement and Proficiency

ABSTRACT

Recent test results reveal that the test score decline has ended, but the legacy of this highly publicized educational episode continues. One widespread interpretation of the decline and recovery is that permissiveness and a collapse of standards in the late 1960s led to the decline and that a return to the basics and to tougher standards caused the turnaround. This paper argues that the causes of test score decline are still uncertain; that schools' standards and programs in the late 1960s probably had less to do with test score decline than is commonly believed; and that although the recent turnaround in test scores may have been the result of schools' renewed emphasis on the skills included on standardized tests, there are risks in teaching to test, in advocating a return to "the basics," and in believing that the central instructional problems of schools today will be remedied by rehabilitating the nation's average scores on standardized tests. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

R. Rossmiller

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Program Report 85-8

THE TEST SCORE DECLINE IS OVER: A REINTERPRETATION

Lawrence C. Stedman and Carl F. Kaestle

Wisconsin Center for Education Research
School of Education
University of Wisconsin
Madison, Wisconsin

October 1985

An abridged version of this paper will appear in Phi Delta Kappan, November 1985.

The research reported in this paper was funded by the Wisconsin Center for Education Research which is supported in part by a grant from the National Institute of Education (Grant No. NIE-G-84-0008). The opinions expressed in this paper do not necessarily reflect the position, policy, or endorsement of the National Institute of Education.

Wisconsin Center for Education Research

MISSION STATEMENT

The mission of the Wisconsin Center for Education Research is to understand, and to help educators deal with, diversity among students. The Center pursues its mission by conducting and synthesizing research, developing strategies and materials, and disseminating knowledge bearing upon the education of individuals and diverse groups of students in elementary and secondary schools. Specifically, the Center investigates

- diversity as a basic fact of human nature, through studies of learning and development
- diversity as a central challenge for educational techniques, through studies of classroom processes
- diversity as a key issue in relations between individuals and institutions, through studies of school processes
- diversity as a fundamental question in American social thought, through studies of social policy related to education

The Wisconsin Center for Education Research is a noninstructional department of the University of Wisconsin-Madison School of Education. The Center is supported primarily with funds from the National Institute of Education.

Table of Contents

	Page
Abstract	vii
When Did It End?	1
What Did Critics Claim About the Test Score Decline?	2
Did the Decline Coincide With Late-1960s Disruptions?	2
How Bad Was It?	3
Who Is Taking the Tests?	4
How Big Was the Decline?	5
What's Wrong With These Particular Measures?	7
Did All Test Scores Decline?	9
What Caused the Decline?	10
Why Did the Decline End?	14
What Do We Do Now?	17
References	19
Sources of Test Score Data	23

ABSTRACT

Recent test results reveal that the test score decline has ended, but the legacy of this highly publicized educational episode continues. One widespread interpretation of the decline and recovery is that permissiveness and a collapse of standards in the late 1960s led to the decline and that a return to the basics and to tougher standards caused the turnaround. This paper argues that the causes of test score decline are still uncertain; that schools' standards and programs in the late 1960s probably had less to do with test score decline than is commonly believed; and that although the recent turnaround in test scores may have been the result of schools' renewed emphasis on the skills included on standardized tests, there are risks in teaching to tests, in advocating a return to "the basics," and in believing that the central instructional problems of schools today will be remedied by rehabilitating the nation's average scores on standardized tests.

Recent renormings of the major standardized tests show that the test score decline has finally ended. On the 1982 Stanford Achievement Tests, for example, eleventh graders scored four percentile points higher in mathematics and ten percentile points higher in reading than their 1973 counterparts. Eighth graders were up six percentile points in math and seven in reading. Students in other grades showed similar improvements, across subjects as diverse as science and spelling. The Iowa Tests of Basic Skills, given to third through eighth graders, also showed general improvement, with scores rising dramatically between 1977 and 1984. Preliminary analyses of the 1984 results, for example, indicate that composite scores are at an all-time high for most grades. The Tests of Achievement and Proficiency also show that high school students have improved their performance in most grades and subjects in recent years. National Assessment of Education Progress (NAEP) results show that thirteen-year-olds' mathematics scores rose between 1978 and 1982; nine- and seventeen-year-olds' scores remained stable ending their previous decline. Seventeen-year-olds' reading scores on the NAEP tests rose between 1980 and 1984.

The score decline on college entrance tests has also bottomed out. American College Test (ACT) scores in English, social studies, and science, for example, have been stable for many years. Although ACT mathematics scores continued to drop until 1983, they recently rose. Scholastic Aptitude Test (SAT) math scores have gone up a few points in recent years. SAT verbal scores have been fluctuating up and down a point or two for several years. The big decline is over.

WHEN DID IT END?

It ended sometime in the late 1970s. The national ITBS results began their upswing in 1977. ACT English scores rose from 1976 to 1978 and have remained roughly the same since. ACT social studies scores have remained stable since around 1976. The SAT decline, in both verbal and math scores, ended around 1980. Results in Iowa, often used as a national barometer because of that state's long history of comprehensive annual testing, show steady increases in most grades since 1978. The President's National Commission on Excellence in Education (1983), with its dire warnings of a nation at risk, was about five years too late. Instead of a "rising tide of mediocrity," it should have proclaimed a rising tide of test scores.

WHAT DID CRITICS CLAIM ABOUT THE TEST SCORE DECLINE?

One widespread interpretation of the test score decline goes like this: there was a massive decline in achievement in the late 1960s and the early 1970s, especially on verbal skills a especially among high school students; it was caused by student-centered permissiveness and is now being cured by a reassertion of tougher standards and basic skills training. (See Harnischfeger & Wiley 1975; Armbruster, 1977; Copperman, 1978; and Brimelow, 1983) This interpretation is full of half truths, and half truths can mislead us in the search for effective educational policy.

Our review of the evidence points to rather different conclusions. Most test scores did not begin falling until the 1970s. This makes it harder to blame the decline on the social movements of the 1960s, though some kind of time-lag theory might be relevant. Furthermore, when it did occur, the test score decline was not as great as portrayed. There was also substantial contradictory evidence throughout the two decades. Some tests, including several of reading and writing skills, did not show declines during the 1970s. In the face of such mixed evidence, we are skeptical that there was a severe, general decline in academic skills. Finally, the educational changes that were purported to have caused the decline were never as widespread as the critics maintained. We believe that some scapegoating has gone on and that the causes of test score decline are still in doubt.

DID THE DECLINE COINCIDE WITH LATE 1960S DISRUPTIONS?

A substantial amount of evidence suggests that test scores particularly at the high school level, remainined stable in the 1960s. The Iowa Tests of Educational Development (ITED), for example, showed steady increases in high school reading and math scores in 1957, 1961, and 1972 renormings. Preliminary Scholastic Aptitude Test (PSAT) renormings showed that high school juniors maintained their reading and math scores in 1960, 1966, and 1974 national comparisons. (There was a slight verbal decline but a slight math increase in the 1966-1974 comparison.) The Metropolitan Achievement Test scores for seventh through ninth graders were equal or higher on reading, math concepts, science, and social studies in 1970 than in 1958. (Math computation performance worsened

for seventh and eighth graders, but improved for ninth graders.) The SAT did show major drops beginning in 1963, but the College Board's Advisory Panel on the SAT decline found that between two-thirds and three-fourths of the 1960s decline was due to the changing composition of test takers. They placed the real skill decline (decline in performance within comparable social groups) primarily in the 1970s.

Their conclusion is supported by renorming studies of most standardized tests, including the California Achievement Tests (CAT), the Metropolitan Achievement Tests (MAT), the ITBS, the Science Research Associates (SRA) test, and the ITED, which showed that the major drops occurred between 1970 and 1978.

HOW BAD WAS IT?

Phrases like "massive decline" (Copperman, 1978, p. 29), and "an almost unremitting fall" (Armbruster, 1977, p. 4) were typical. Paul Copperman (1978) called it the "first major skills decline in American educational history" (p. 39). Critics presented test score data in various statistical guises, many of them quite dramatic. Copperman (1979) argued that the average (50th percentile) high school student of the late 1970s ranked at only the 39th percentile of his 1965 counterparts. The SAT verbal drop was almost one-half of a standard deviation, a big shift in the distribution of scores. Several tests showed that eleventh and twelfth graders lost a year or more in measured reading ability during the 1970s. (Bode, 1981b, p. 4; CTB/McGraw-Hill, 1974, n.d.) If one focuses on the SATs and on standard deviations, the test score declines do appear substantial, but there are several problems with these stark descriptions.

Who Is Taking the Tests?

First, the figures are unadjusted for changes in the composition of test takers. We must distinguish decreases due to a widened or changing testing pool from those due to a general decline in students' skills. Trends on college entrance exams such as the SAT are especially difficult to interpret because they measure the performance of a self-selected group of students whose composition changes annually. Even in the 1970s, changing composition accounted for a substantial portion of their

decline. More students from characteristically lower-scoring groups continued to take college entrance tests, including minority students, those intending to pursue "career" majors as opposed to "arts and sciences" majors, women (who score lower, on average, in math), and students going to two-year community colleges and four-year public universities as opposed to highly selective liberal arts colleges. The College Board's Advisory Panel estimated that between 20 and 30 percent of the SAT decline in the 1970s was still due to such changes (Advisory Panel, 1977).

Changes in family size also had an impact, although not as great as some once claimed. First- and second-born children score higher than average, later-borns score lower. In a background report for the Advisory Panel, Breland (1977) estimated that such changes accounted for about 16 percent of the verbal SAT decline between 1964 and 1976. A recent study of the period from 1971 to 1977 produced a 4 to 9.4 percent estimate, although the effect could run higher. (Zajonc & Bargh, 1980)

Combining the estimated birth-order effects and estimated effects of changing characteristics of test takers means that at least 24 to 40 percent of the 1970s SAT decline was not due to changes in the schools' effectiveness in skill training. Many other nonschool factors may also have contributed, as we shall discuss.

Although reported trends on standardized test batteries like the CAT and the ITED are an improvement over reported trends on SAT tests in that they are based on nationally representative samples, they too can be affected by nationwide changes in immigration, dropouts, or birth order. These effects may have been as substantial in the 1970s as those affecting the SAT. falling black dropout rate and increased Asian and Hispanic immigration increased the percentage of minority students in our high schools from one-sixth to nearly one-fourth (National Center for Education Statistics, 1979, p. 17; Bureau of the Census, 1981a, p. 35). Such changes likely contributed to lowered scores. Birth-order effects also contributed, and, for unknown reasons, are greater on standardized achievement batteries than on the SAT (Zajonc & Bargh, 1980). Another contributing factor, albeit small, was the changing age of students. Due to earlier school entering ages and more automatic promotion policies, students coming into a given grade were increasingly younger. Researchers who have studied

long-term test score trends have stressed the necessity of accounting for the differential maturity of students (Gates, 1961; Farr, Fay & Negley, 1978). Adding this factor to the birth-order and composition factors mentioned above suggests that demographic changes may account for between 30 and 50 percent of the 1970s achievement test score decline at the high school level. Critics, however, tend to assume that virtually all of the 1970s declines were due to instructional failure (Brimelow, 1983; Copperman, 1978, 1979; Ravitch, 1985, 172-181).

Copperman, for example, presumed that the skill decline began in the mid-1960s and dropped steadily thereafter, thus ignoring a huge compositional effect. His claim of a drop across achievement tests to the 39th percentile was based on an estimated 2.5 percent of a standard deviation drop per year from 1965 to 1978. Thirteen years of such a drop yielded an overall 32 percent drop, or 11 percentile points. Since this figure was unadjusted for compositional effects, and since the real skill decline primarily occurred in the 1970s, the overall decline was much less than Copperman claimed. Figuring 1.3 to 1.8 percent of an SD per year for seven years (the 1970's decline minus the estimated compositional effect) produces a total decline of 9.1 to 12.6 percent of a standard deviation during the 1970s. This amounts to a drop of only four to six percentile points, to the 44th or 46th percentile level. Similarly, adjusting grade-level scores for compositional effects reduces apparent declines by up to one half. Thus, on tests in subjects that showed as much as a whole year decline, the adjusted score would be a half year.

How Big Was the Decline?

A second problem is that critics rarely relate test score declines to actual skills. What was the difference in skills between students who scored one-half a grade level lower than another earlier group? What specific tasks could students no longer do? Standardized tests are constructed in such a way that small shifts in test performance produce large changes in percentile and grade equivalent rankings. The decline thus sounds large when described in grade equivalent and percentile terms, even though the actual performance drop could be quite small. Oscar Buros, the late editor of the Mental Measurements Yearbook, argued against the reliance on normed scores for

interpreting educational achievement. He believed that grade level equivalents give people a vague and misleading impression of skill levels. He advocated getting closer to actual performance by using the percent of items a student knows (Buros, 1978, p. 1976). If we describe student performance in this way, we get a different sense of the magnitude of a skill decline. On many standardized tests, differences between grades amount to only a few percentage points, particularly at the high school level. On the SRA, ninth through twelfth graders' reading scores dropped a half to a full grade level between 1971 and 1978, but this corresponded to a small drop in the percent of items answered correctly. Twelfth graders, for example, had dropped a whole grade level in reading, but this was only from 72 to 68 percent correct, a four percent drop. Mathematics declines were similar (Bode, 1981b, p. 4; Bode, 1981a, p. 33). Furthermore, these figures are unadjusted for compositional changes, so the actual skill decline among similar students was smaller yet. Several of the NAEP tests showed declines, but these also showed only small drops in performance. Between 1970 and 1980, for example, in inferential reading comprehension, seventeen-year-olds dropped from 64 to 62 percent, a 2 percent drop; thirteen-year-olds went from 56.1 to 55.5 percent, or only a .6 percent drop. In math, from 1973 to 1982, seventeen-year-olds dropped from 52 percent to 48 percent correct, or only 4 percentage points, while thirteen-year-olds dropped only 2 points. In science, from 1970 to 1977, seventeen-year-olds dropped only 4.7 percentage points, thirteen-year-olds 2.4 points. Other tests may show larger declines, but the point is the same: when they are expressed in terms of percent correct, they do not seem as great as when they are expressed in grade levels.

Another way of assessing the decline is to ask at what percentage of their former skill levels are students now performing? On the NAEP tests, for example, students were performing at 97 percent of their former levels in inferential comprehension, 92 percent in mathematics. High school students on the SRA were reading at about 95 percent of their former levels. Some may believe that even a five percent decline in skill level is worrisome. The Nation At Risk report argued that such skill declines threatened our very economic security as a nation. But what are the demonstrable educational and economic ramifications of test score declines?

In fact, the statistical links between academic success at one level and the next are relatively weak, as are those between academic performance and economic performance. The correlation between SAT scores and freshmen grades, for example, is about .40 (Advisory Panel, 1977, p. 9). The decline of 20 percent of a standard deviation in SAT scores in the 1970s (which accounts for compositional changes) would translate into a drop of only ten percent of a standard deviation in freshman grades. The correlation between achievement test scores and measures of job proficiency is around .25 (Olneck, 1984, citing Schmidt & Hunter), so the drop of 12.6 percent of a standard deviation in high school standardized achievement test scores during the 1970s would translate into a drop in job performance of only 3.1 percent of a standard deviation. Furthermore, as Olneck pointed out, new workers comprise only a small proportion of the entire workforce, so recent declines in productivity can hardly be linked to recent changes in test scores. Even ten years of reduced skills among all new workers would only affect about one-fourth of the active work force. The combined effect of the modest correlations makes the critics' attempts to link declining test scores with changes in industrial productivity downright silly.

All of the above discussion presumes that the tests are completely valid measures of academic skills. Yet performance on the tests reflects an undetermined proportion of other factors, such as motivation and test-taking skills. If these extraneous factors could be accounted for the actual skill decline that occurred during the 1970s would have been even smaller than described above.

What's Wrong With These Particular Measures?

A third set of problems arises from the deficiencies of the particular measures most often used as evidence of a skill decline: college entrance exams, national achievement test renormings, trends in individual states' achievement test scores, and the NAEP tests.

College entrance exams provide an annual barometer of performance changes but, as noted above, the composition of the test-takers changes annually, and thus it is imperative to adjust for the compositional effect. They are further limited as a national barometer because they

apply primarily to the college-bound student rather than the average student.

Standardized achievement test trends are derived from periodic renormings (five to seven years apart, generally) carried on by publishers when redesigned tests are introduced. The new test is given to a nationally representative sample of schools. Performance on the new test, and thus current national performance, is linked to old results through "equating studies," in which samples of contemporary students are given both the new and the old test. Problems with equating abound. Often two different contemporary groups are given the different test versions; sometimes only portions of the two test versions are administered. Also, the equating samples are usually not representative of the nation, often involving a few school districts or a small fraction of the national norming sample. Even some test publishers warn against the use of renorming data to infer national trends. Metropolitan Achievement Test publishers stated in 1978 that "these data are not appropriate for making generalizations concerning changes over time in the relative achievement of American students in the basic skills areas" (The Psychological Corporation, 1978, p. 1). More recently, these same publishers warned that there is a "popular misconception about changing norms: that a change in the norms from an old test to a new test reflects a change in the ability of the reference groups over time." On the contrary, "there are simply too many complex and confounding variables to make a sound judgment about performance over time" (Test Department, 1983, pp. 1,2). They cite changes in the national samples of students and the changing relevance of the test content as factors that confound any generalizations. Critics who used renorming evidence to describe national trends typically do not discuss these serious limitations.

State trends on achievement tests are problematic because the data for the 1960s and 1970s were limited to a handful of states. Iowa is often used as a barometer, but it can hardly be considered representative. It is predominately rural, and has few minority students, and thus performance there can hardly be said to reflect the nation. Even among similar states, the trends are ambiguous; for each state the critics cited, showing declines for the 1960s and early 1970s, there was a matching state that did not. Alabama and South Dakota high school scores were stable in the 1960s, for example; Mississippi eighth graders and Michigan seventh graders

had stable scores in the early 1970s (See Armbruster, 1977; Farr, Tuinman & Rowls, 1974).

The NAEP tests are probably the best indicators of national trends. They are drawn from nationally representative samples like the standardized tests, but trends on common items are reported so there is no problem relating results from old and new versions of the tests; renorming studies are unnecessary. Furthermore, NAEP results, unlike results from standardized tests, are regularly broken down and reported by racial, geographical, and SES groups, so that trends by subgroups can be followed. Test items are also made public, so schools can independently examine what kind of skill is being measured. They are limited like the other tests, however, in that the items on the test may not reflect what is taught in schools, or may test only a small portion of it.

Did All Test Scores Decline?

A fourth problem with the argument of the test decline alarmists is that they paid little attention to the contradictory evidence of the 1970s. The National Assessment of Educational Progress showed that thirteen- and seventeen-year-olds maintained their overall reading scores and nine-year-olds improved theirs in 1970, 1975, and 1980 testings. Seventeen-year-olds slipped in inferential comprehension but, as noted, the drop was minor--from a 1970 level of 64 percent correct to 62 percent in 1980. Furthermore, this decline was not universal. The only region experiencing statistically significant declines in inferential skills was the Northeast; boys showed such declines, but not girls. Blacks' scores did not fall off significantly.

Some commentators argue that the NAEP reading tests are easier than standardized high school achievement tests and test lower-level skills. In fact, the percent of questions missed by seventeen-year-olds on NAEP is comparable to that on other achievement tests, and the proportion of the test devoted to to inferential skills is also similar.

In functional literacy NAEP results showed that seventeen-year-olds improved their performance from 1971 to 1975. In writing, the NAEP showed that seventeen-year-olds' skills remained roughly the same between 1969

and 1979. Rhetorical skill on narrative tasks rose during the period, as did cohesion scores. A comparison of results on the Metropolitan Achievement Tests and the Stanford Achievement Tests showed a five to six month gain in reading and a six to twelve month gain in math for grades seven through ten from 1973 to 1978. The ACT natural science scores have remained stable over the past two decades. Finally, those who cite the SATs as evidence for the decline rarely mention data from ETS's Achievement Tests. Scores in English composition, biology, chemistry, physics, French, and Spanish showed increases from 1967 to 1976, the time of the worst SAT decline. Thus, although the students who took these tests in 1976 had lower SAT scores than their predecessors, they outscored them on the achievement tests. (Advisory Panel, 1977, p. 22) Finally, Flynn (1984) reviewed sketchy evidence to suggest that IQ scores were stable or rising through the period 1972 to 1978, during the steepest of the SAT decline. The evidence about a massive, consistent skill decline, then, is much more mixed than the achievement critics claim, and the contradictory evidence is not easily explained.

WHAT CAUSED THE DECLINE?

Setting aside the contradictory data for the moment, what about the causal explanations for the alleged declines in academic achievement indicators?

Frank Armbruster, of the Hudson Institute, claimed that during the 1960s " . . . acceptance of improper behavior, and even some types of criminal acts, were becoming commonplace. Adults, even police, could be ignored with impunity" (1977, p. 8-9). He further claimed that "moderates in our school system lost their prominence and some apparently injudicious activist educators gained influence" (p. 7). These activists allegedly altered curriculum and teaching methods and opened the schools to the values of the slums (p. 8). Teachers yielded to students the "responsibility of determining when, if, and within a disturbingly questionable range, even what they would study" (p. 9). When achievement declined, sympathetic media and school boards let them fix blame anywhere but on the schools. "This sympathetic attitude," according to Armbruster, "may have been a 'spinoff' from the Kennedy era and later emphasis on the 'War on Poverty'" (p. 8).

Paul Copperman (1978) blamed the decline to a great extent on the open education movement and a breakdown in authority relations. He criticized the "undisciplined counter-culture approach recommended by Kohl and others of his ilk" (p. 64). He described Silberman's Crisis in the Classroom, which advocated more freedom and openess in education, as "one of the most damaging pieces of educational writing to have been published in the past twenty years," and claimed that Silberman's recommendations were widely adopted (p. 68). Like Armbruster, Copperman claimed that "it is current educational policy to give children a great deal of choice over what, how, and even whether they study. In every school in the country, students determine how much science, math, history, and composition they will take" (p. 150). He also argued that free health clinics, runaway centers, and alternative high schools had undermined educational authority by convincing young people that "irresponsibility, hedonism, and laziness comprise an acceptable alternative value system . . . " (p. 170). Echoes of these late 1970s conservative complaints persist in the mid-1980s.

There are several problems with these explanations. First, they are overstated. If we were to imagine for a moment that everything happened the way Armbruster and Copperman claimed, then why, in the face of such a major social and educational breakdown, was there only a decline in some types of test scores and evidence from other tests of stability or improvement? The results suggest more educational resilience than the critics recognize and less social deterioration. Second, the timing of their explanations is off. Since the decline actually occurred in the 1970s, blaming it on the social movements of the 1960s requires a more subtle lag theory. We do not agree that widespread unrest and disobedience affected most high schools across the nation. But even if we did, the years of greatest student protest occurred between 1968 and 1971, while the decline in standardized test scores took place from 1971 to 1978, and SAT scores continued to slump until 1980. A student who was in twelfth grade in 1978 would not have been in high school during the protest years, but would instead have been proceeding from the second through the fifth grade. His or her high school years would have been from 1975 to 1978, hardly a time of protests or educational experimentation. Blaming the decline on the schools may be fashionable, but the middle and late 1970s were a time of educational retrenchment with a renewed emphasis on the basics, the spread of

statewide competency testing, and actions to end social promotions. We can hardly blame the test declines of the 1970s on activist educators who, frustrated at their inability to change the schools, had effectively abandoned their efforts by the mid-1970s. An interesting variant on the 1960s thesis is proposed by Jencks (1980) who placed major emphasis on teachers' loss of nerve and retreat from authority as an explanation for test score decline. We believe that there is some truth to the notion that the Vietnam - Watergate years led to some disillusionment with rationality and tradition and that this may have had some lingering effect in the public schools. But the pervasiveness of this sentiment has not been established; nor do we look upon this reaction among youth and educators as unjustified or wholly undesirable. To decide that flexibility, student activism, and even a touch of rebelliousness are harmful because they detract from the pursuit of the skills measured by standardized tests would be short-sighted. Expanded opportunities for student initiative and creativity can enhance intellectual growth and motivation for learning. Whatever one's philosophical views, the loss-of-nerve explanation suffers from bad timing and weak evidence; it also lends itself to scapegoating.

Nor can we accept the explanations of the President's National Commission on Excellence, which claimed that the high school curriculum had been "homogenized, diluted, and diffused" and that the resulting "curricular smorgasbord . . . explains a great deal about where we find ourselves today" (p. 18). Their primary evidence was a flawed study of high school transcripts comparing student records from 27 high schools in the late 1960s with a national sample in the late 1970s. The two groups were not comparable, and the differences in academic courses were small and sometimes favored the later group (see Stedman and Smith, 1983). A more recent study that involved nationally representative samples showed increases, not decreases, in academic course enrollments through the 1970s (West, Diodato, & Sandberg, 1984). Blaming achievement test score declines on a relaxation of academic standards is too facile an explanation for the mid-1970s, when many contradictory pressures affected the nation's schools. Some research belies the causal connection. Echternacht (1977) studied two groups of high schools, those that maintained their SAT scores and those with declines worse than the national decline, and found they did not differ in their educational approach. Truancy, discipline problems, and teacher permissiveness

had increased similarly. The differences in the number of academic courses taken by students were tiny, sometimes favoring test-score decliners. English curricula were similar; pass-fail grading and nontraditional offerings had expanded to the same extent.

At the elementary school level, blaming the decline on the open classroom movement also misses the mark. Although the timing is right, the movement never was as widespread as Copperman believes. In 1970, after several years of visiting schools across the country, Silberman complained about "what grim, joyless places most American schools are, how oppressive and petty are the rules by which they are governed, how intellectually sterile and esthetically barren the atmosphere . . . " (1970, p. 10). He found few examples of open schools. By 1973, Silberman had observed a shift in educational attitudes and practice. Some teachers and schools had adopted open classrooms, but he added: "the classrooms I am talking about constitute a small handful" (Silberman, 1973, p. xvii). John Holt, another open classroom advocate, observed in 1976 that "many of these innovations were dying or dead soon after [Silberman's] book came out . . . . Most of these lasted only a few years." He concluded that "there never was much 'open education.'" Citing a study of Minnesota schools, he suggested that less than one percent of the nation's children were in open classrooms (pp. 140, 143-4).

Whatever its extent, the adoption of open classrooms did not harm national scores. On the NAEP test, for example, nine-year-olds improved their reading scores from 1970 to 1975. Even the educational critics agreed that elementary school students had stable or rising scores during the 1970s. Furthermore, the American Institutes for Research, reporting a study of 30,000 elementary and junior high school students in thirteen school districts in nine states, found that their achievement was not related to the level of educational innovation. (Advisory Panel, 1977, p. 41).

Given our view that the actual skill decline was less drastic than is typically claimed, a series of nonschool factors, combined or separately, could account for a large part of it: drugs, television, extracurricular activities, decreases in test motivation, etc. Insufficient attention has been given to the social and economic disruptions of family life which increased during the 1970s and may have hurt achievement. From 1970 to 1978, the divorce rate

rose from 3.5 per 1000 to 5.2; one parent families increased; white female headed households increased from 9 to 11.5 percent; black ones from 27 to 36 percent. The unemployment rate in the first half of the 1970s jumped from 4.9 to 8.5 percent (Bureau of the Census, 1981b). Millions gave up looking for work. We believe such upheavals in the families of junior or senior high school students likely contributed to lowering their achievement.

WHY DID THE DECLINE END?

"Failure is an orphan, but success has a thousand parents." Writing in 1976, Tavris predicted that when test scores started to rise again everyone would take credit.

> City X will attribute the upswing to their open schools, city Y will praise their unflinching attack on permissiveness. Educator A will pat himself on the back for his summer-school program, and educator B will pass the champagne to everyone involved in her "back-to-the basics" approach. (p. 74)

This is indeed happening. Governors, state legislators, and the sponsors of the recent reform reports have suggested that their actions were important in turning things around. In February 1985, President Reagan told the National Association of Secondary School Principals that both aptitude and achievement tests "underwent a virtually unbroken decline" from 1963 until 1981, but that "since our administration put education at the top of the American agenda we've seen a grassroots revolution" to get back to "basic teaching and learning" (1985a, Feb. 7, p. 2). Later that month he reiterated to a group of educators that SAT scores started to climb after 1980 because he started education on "the long, hard road to excellence" (Robinson, 1985). Yet achievement test scores were already rising well before President Reagan's election, before reformers issued their reports, and before legislatures passed their post-1980 reform bills.

We attribute a portion of the reversal to the changed family configuration of test takers, who are now more often first or second born. Zajonc, the major proponent of birth order explanations, actually predicted in the mid-1970s that the test score decline would bottom out in 1978 and that scores would then start rising (Tavris, 1976). We also believe an increasing drop-out rate, which

makes the high school compositon more selective, contributes to rising scores. Changes in other nonschool factors may also have produced part of the increase.

Changes in the schools, of course, are a factor in the test score turnaround. During recent years, instruction has focused more on the skills measured by standardized tests. Many states and individual school districts adopted competency tests for promotion and graduation. Effective schools research, with its emphasis on the frequent and systematic testing of pupils, has influenced educational reform. Testing is central to school improvement projects throughout the country. Criterion-referenced testing, which directly links curriculum objectives with test items, has also assumed a greater role. With all this increased attention to tests, it is hardly surprising that students have improved their test scores.

Do the improved test scores of the past seven or eight years signal improvement in the schools?

We doubt that the current focus on testing and traditional pedagogy can solve the two of our schools' most pressing educational problems. First, schools have had a continuing problem teaching higher order skills to a majority of their students. Thus, to blame our educational ills on open classrooms and believe that back-to-the-basics programs can cure them is illogical. It may also be counterproductive. Borkow (1982) suggested that the schools are "hitting the basics too hard," and cited evidence that this emphasis has hurt the development of higher-order skills and brought down the test scores of the higher-achieving students. Consultants for the NAEP mathematics program blamed the 1970s decline in problem solving, in part, on the back-to-the-basics movement (NAEP, 1979, p. 25).

Second, for many years, the tests have revealed a large number of students deficient in the literacy and computational skills needed by workers and citizens. Since the pool of students who lacked fundamental skills was large even fifteen years ago, returning to the much vaunted instructional strategies of the past isn't a logical solution to this problem either. Results from the NAEP tests illustrate this point. Consider the following two mathematical application items. Students were given an electric bill with a charge of $9.09 for 606 kilowatt hours of consumption. The question was: what is the cost

per kilowatt hour? The percentage of seventeen-year-olds answering this correctly declined seven percentage points from 1973 to 1978. More notable is the fact that 88 percent of the seventeen-year-olds couldn't answer the problem in 1973 even though they had gone to elementary and junior high school in the good old days (NAEP, 1979, p. 12). In the second item, a hockey team won five of its twenty games. What percentage of the games did it win? On this question, scores were down eight percentage points, but the decline pales when compared to the fact that at the high point 40 percent of the students could not answer it correctly. Going back further in time, to the 1964 international comparison in math achievement, U.S. thirteen-year-olds did quite poorly compared to students in other industrialized nations, even though the bulk of their schooling took place in the 1950s and early 1960s, the supposed heyday of U.S. education. (Husen, 1967)

Tests of reading comprehension show similar results. In the NAEP studies (1981, 1982), for example, few items showed declines, but on many, the percentage of students who couldn't answer correctly in either 1970 or in 1980 was alarming. On several basic reading questions, from one-fourth to well over one-half of the students answered incorrectly. Similar results were found in the Mini Assessment of Functional Literacy (Gadway & Wilson, 1976).

All the talk about test score declines and getting back to the basics, therefore, tends to obscure the long-standing failure of the schools to reach the lower third of our students effectively. They need basic skills, but back-to-the-basics pedagogy, as popularly understood--more discipline, tougher grades, and traditional textbooks--may serve the low achievers no better in the late 1980s than it did in the early 1960s.

Changed attitudes among students and teachers, tougher high school graduation requirements, more homework, and more testing may improve some students' tests scores, but we must be careful that in erecting tougher standards we do not lose other students. Higher drop-out rates, along with reduced aid to low-income college-bound students, may raise achievement test scores, but it's nothing to be proud of. Our goal should be to improve all students' academic skills.

The recent focus on testing is already having serious negative consequences. As a result of the effective

schools movement, Cuban (1983) found that many school systems are returning to the pre-1900 notion of a uniform curriculum, using a single set of textbooks for a given grade, regardless of individual differences. Many teahers are returning to the old-fashioned whole-group method, combining lecture, recitation, and seatwork, a strategy which Cuban believes goes "far beyond what the research suggests". He criticizes the "single-minded quest for higher test scores" for narrowing the schools' agenda to that which is easily measured. Less attention is being given to other goals that educators and parents value, such as sharing, learning to make decisions, developing self-esteem, and acquiring higher level thinking skills and aesthetic sense. Daniel and Lauren Resnick present evidence from Pittsburgh to show that minimum competency testing encourages teachers to focus only on the minimum skills to be tested (Resnick & Resnick, 1985, p. 15). Meier (1981, 1984), who has taught in inner-city schools for the past two decades, finds the focus on te-ting is harming the development of good reading skills. The quest for high test scores has had particularly bad effects on children from poor families. In schools attended mainly by low-income children, Meier notes "the prevalence of programmed scripts based on behavior-mod techniques, reading "kits" consisting of hundreds of unrelated paragraphs followed by multiple-choice questions and reams of ditto sheets." Worse, Meier noted, "lower-class schools are often devoid of books (except perhaps workbooks, readers, and the textbook); instead of libraries they have remedial reading and audiovisual 'labs.' It's not universal, but it's common" (p. 63).

Such problems may worsen and spread as the pressure for accountability grows and test score rehabilitation is made the central educational goal. We question, therefore, the wisdom of making the test score decline the focus of educational reform efforts.

WHAT DO WE DO NOW?

Standardized tests reveal national trends, and the problems are a national concern. But the solutions must be largely local. Schools in different settings have different problems. Educators and school boards will have to identify their students' most pressing needs, ranging from rudimentary skills to critical thinking and from positive self-image to positive attitudes about academic work. Do students like to read? Can they apply their

math skills? Are they achieving in areas not easily tested? Are they learning to appreciate and respect other cultures? Is the curriculum balanced, fair, engaging, inclusive, demanding?

Tests have played an exaggerated role in recent discussions of educational reform. There was, no doubt, some decline in the type of skills measured by standardized tests in the 1970s, and there have always been too many children lacking rudimentary skills. Therefore we applaud the renewed emphasis on reading and writing and the revived efforts to provide good academic training in high schools. But we dissent from the perspective that blames a collapse of standards on the turbulent 1960s and looks forward to rising test scores as signals that our educational problems have been solved.

The challenge is not to get kids back into harness and crack the whip. If that were a good solution, it might be simple. But unfortunately, schools must not only improve basic and higher-level skills but also become better places for teachers to work and for children to learn about themselves and their society. This cannot be done by top-down, test-based solutions. The challenge, then, is for each community to find a philosophically appealing and educationally effective balance between common experiences and cultural diversity, between a supportive atmosphere and standards of excellence, between student initiative and the transmission of uplifting knowledge. In this process tests can play only a limited role.

REFERENCES

Advisory Panel on the Scholastic Aptitude Test Score Decline (1977). On further examination. New York: College Entrance Examination Board.

Armbruster, F. (1977). Our children's crippled future: How American education has failed. New York: Quadrangle Books.

Bode, R. K. (1981a). SRA achievement series: Technical report number 2. Chicago: Science Research Associates.

Bode, R. K. (1981b). SRA achievement series: Technical report number 3. Chicago: Science Research Associates.

Borkow, N. (1982). Analysis of test score trends: Implications for secondary school policy--A caution to secondary school administrators. Washington, D.C.: National Institute of Education.

Breland, H. M. (1977). Family configuration effects and the decline in college admissions test scores: A review of the Zajonc Hypothesis. In Appendix to Advisory Panel (1977), On further examination. New York: College Entrance Examination Board.

Brimelow, P. (1983). What to do about America's schools. Fortune, September 19: pp. 60-64.

Bureau of the Census. (1981a). School enrollment, P-20, Number 374. Washington, D.C.: Government Printing Office.

Bureau of the Census. (1981b). U.S.A. statistics in brief, 1981. Washington, D.C.: Government Printing Office.

Buros, O. (1978). Fifty years in testing. In O. Buros, ed., The eighth mental measurements yearbook. Highland Park, N.J.: The Gryphon Press.

Carver, R. P. (1972). Reading tests in 1970 versus 1980: Psychometric versus edumetric. The Reading Teacher 26: p. 299-302.

Copperman, P. (1978). The literacy hoax: The decline of reading, writing, and learning in the public schools and what we can do about it. New York: William Morrow.

Copperman, P. (1979). The achievement decine of the 1970s. Phi Delta Kappan. June: pp. 736-39.

CTB/McGraw-Hill. (n.d.) CAT A-C equating tables. Monterey, CA: CTB/McGraw-Hill.

CTB/McGraw-Hill. (1974). Technical report: All levels, California achievement tests, 1970 edition. Monterey, CA: CTB/McGraw-Hill.

Cuban, L. (1983). Effective schools: A friendly but cautionary note. Phi Delta Kappan June: pp. 695-6.

Echternacht, G. J. (1977). A comparative study of secondary schools with different score patterns. In appendix to Advisory Panel (1977), On further examination. New York: College Entrance Examination Board.

Farr, R., Fay, L. and Negley, H. (1978). Then and now: Reading achievement in Indiana (1944-45 and 1976). Bloomington: Indiana University. (ED 158 262)

Farr, R., Tuinman, J. and Rowls, M. (1974). Reading achievement in the United States: Then and now. Bloomington: Indiana University. (ED 109 595)

Flynn, J. R. (1984). The mean IQ of Americans: Massive gains, 1932 to 1978. Psychological Bulletin 95: 29-51.

Gadway, C. J. and Wilson, H. A. (1976). Functional literacy: Basic reading performance. Denver: Education Commission of the States. (ED 112 350)

Gates, A. I. (1961). Reading attainment in elementary schools: 1957 and 1937. New York: Teachers College.

Harnischfeger, A., and Wiley, D. (1975). Achievement test score decline: Do we need to worry? St. Louis: CEMREL.

Hoffman, B. (1962). The tyranny of testing. New York: Collier Books.

Holt, J. (1976). Instead of education. New York: Dutton.

Husen, T. (1967). International study of achievement in mathematics: A comparison between twelve countries. New York: Wiley.

Jencks, C. (1980). Declining test scores: An assessment of six alternative explanations. Sociological Spectrum, Premier Issue, December: 1-15.

Meier, D. (1981). Why reading tests don't test reading. Dissent Fall: 457-466.

Meier, D. (1984). "Getting tough" in the schools. Dissent Winter: 61-70.

National Assessment of Educational Progress. (1979). Changes in mathematical achievement, 1973-78. Report Number 09-MA-01. Denver: Education Commission of the States.

National Assessment of Educational Progress. (1981). Three national assessments of reading: Changes in performance, 1970-1980. Report Number 11-R-01. Denver: Education Commission of the States.

National Assessment of Educational Progress. (1982). The reading comprehension of American youth. Report Number 11-R-02. Denver: Education Commission of the States.

National Center for Education Statistics. (1979). The condition of education. Washington: Government Printing Office.

National Commission on Excellence in Education. (1983). A nation at risk: The imperative for educational reform. Washington: Government Printing Office.

Olneck, M. (1984). Terman Redux. Contemporary Education Review 3: 297-314.

Ravitch, D. (1985). The schools we deserve. New York: Basic Books.

Reagan, Ronald. (1985a). Remarks of the President to the National Association of Secondary School Principals 68th annual convention. Press Release. Washington: The White House, February 7.

Resnick, D., and Resnick, L. (1985). Standards, curriculum and performance: Historical and comparative perspective. Education Researcher 14,4: 5-20

Robinson, Thomas. (1985). "President Reagan sees education as important and improving, says students should feel 'reverence.'" Education times, March 4, 1985, p. 1.

Silberman, C. (1970). Crisis in the classroom. New York: Vintage Books.

Silberman, C. ed. (1973). The open classroom reader. New York: Vintage Books.

Stedman, L., and Smith, M. (1983). Recent reform proposals for American education. Contemporary Education Review 2: 85-104.

Tavris, C. (1976). The end of the IQ slump. Psychology Today April: 69-74.

Test Department, Harcourt Brace Jovanovich, Inc. (1983). Some comments on the relationship between scores on the 1973 and 1982 edition of standard achievement tests. Special Report Number 4A. New York: Harcourt Brace Jovanovich, Inc.

The Psychological Corporation. (1978). Revised. Equivalent grade equivalent scores for metro '72 and metro '78. Special Report Number 20. New York: Harcourt Brace Jovanovich, Inc.

West, J., Diodato, L., and Sandberg, N. (1984). Trend study of high school offerings and enrollments; 1972-73 and 1981-82. Washington, D.C.: U.S. Government Printing Office. (ED 253 530)

Zajonc, R. B., and Bargh, J. (180). Birth order, family size, and decline of SAT scores. American Psychologist 35: 662-668.

SOURCES OF TEST SCORE DATA

American College Test (ACT):

American College Test. (1985). Table. ACT score means & SD's for successive years of ACT-tested college-bound students. Iowa City: American College Test.

Advisory Panel on the Scholastic Aptitude Test Score Decline. (1977). On further examination. New York: College Entrance Examination Board. Page 21.

California Achievement Tests (CAT):

CTB/McGraw-Hill. (1974). Technical report: All levels, California Achievement Tests, 1970 edition. Monterey, CA: CTB/McGraw-Hill.

CTB/McGraw-Hill. (1979). California achievement tests: Forms C and D level, 10-19, technical bulletin 1. Monterey, CA: CTB/McGraw-Hill.

CTB/McGraw-Hill. (n.d.) CAT A-C equating tables. Monterey, CA: CTB/McGraw-Hill.

Comprehensive Tests of Basic Skills:

Harnischfeger, A., and Wiley, D. (1976). Achievement test score decline: Do we need to worry? St. Louis: CEMREL.

CTB/McGraw-Hill (1982). Comprehensive tests of basic skills: Preliminary technical report forms U and V. Monterey, CA: CTB/McGraw-Hill.

CTB/McGraw-Hill (1982). Equating tables for CTBS U and V and CTBS S and T. Monterey, CA: CTB/McGraw-Hill.

Iowa Tests of Basic Skills:

Hieronymus, A. N. (1985). Preliminary information on the 1984 ITBS results. Personal Communication, September 16, 1985.

Hieronymus, A. N., Lindquist, E. F., and Hoover, H. D. (1982). Manual for school administrators: Iowa Tests of Basic Skills. Chicago: Riverside Publishing Company.

Riverside Publishing Company. (1983). The development of the 1982 norms for the Iowa Tests of Basic Skills, Cognitive abilities test, and tests of achievement and proficiency. Chicago: Riverside Publishing Company.

Iowa Testing Program. Summary of national achievement trends, 1955-1981, grades 3 - 8, using median performance of 1955 fall as a base. Mimeo provided by A. N. Hieronymus.

Iowa Tests of Educational Development:

The Iowa Tests of Educational Development: A summary of changes in the ITED norms. Iowa City: The University of Iowa, 1971.

Science Research Associates. (1978). Iowa Tests of Educational Development: Technical report, 1978. Table XIV, p. 13.

Metropolitan Achievement Tests (MAT):

Test Department, Harcourt Brace Jovanovich, Inc. (1971). Equivalent scores for Metropolitan Achievement Tests, 1970 edition, and Metropolitan Achievement Tests 1958 edition. Special Report Number 14. New York: Harcourt Brace Jovanovich, Inc.

The Psychological Corporation. (1978). Equivalent grade equivalent scores for metro '70 and metro '78. Special Report Number 20, Revised. New York: Harcourt Brace Jovanovich, Inc.

The Psychological Corporation. (1978). Equivalent grade equivalent scores between metro '78 and the 1973 Stanford Achievement Test. Special Report Number 21. New York: Harcourt Brace Jovanovich, Inc.

National Assessment of Education Progress (NAEP):

Gadway, C. J., and Wilson, H. A. (1976). Functional literacy: Basic reading performance. Denver: Education Commission of the States. (ED 112 350)

National Assessment of Educational Progress. (1981). Three national assessments of reading. Changes in achievement, 1970-1980. Report Number 11-R-01. Denver: Education Commission of the States.

National Assessment of Educational Progress. (1979). Changes in mathematical achievement, 1973-1978. Report Number 09-MA-01. Denver: Education Commission of the States.

National Assessment of Educational Progress. (1981). Three national assessments of science. Changes in achievement, 1969-1979. Report Number 08-S-35. Denver: Education Commission of the States.

National Assessment of Educational Progress. (1982). Writing achievement, 1969-1979. Report Number 10-W-35. Denver: Education Commission of the States.

National Assessment of Educational Progress. (1983). The third national mathematics assessment: Results, trends and issues. Report Number 13-MA-01. Denver: Education Commission of the States. Preliminary Scholastic Aptitude Test:

Breland, H. M. (1977). The SAT score decline: A summary of related research. Appendix to Advisory Panel (1977), On further examination. p. 25.

Scholastic Aptitude Test (SAT):

Harnischfeger, A, and Wiley, D. E. (1975). Achievement test score decline: Do we need to worry? St. Louis: CEMREL.

Admissions Testing Program of the College Board. (1983). National report on college-bound seniors, 1983. New York: College Entrance Examination Board.

Advisory Panel on the Scholastic Aptitude Test Score Decline. (1977). On further examination. New York: College Entrance Examination Board.

Science Research Associates Tests (SRA):

Bode, R. K., (1981). SRA achievement series: Technical report Number 3. Chicago: Science Research Associates, Inc.

Cleary, T. A., and McCandless, S. A. (1976). Summary of score changes (in other tests). Appendix to Advisory Panel (1977), On further examination. New York: College Entrance Examination Board.

Science Research Associates. (1980). Test talk, 80-2: Comparing the results of testing with the 1971 and 1978 editions of the SRA achievement series. Chicago: Science Research Associates.

Sequential Tests of Educational Progress:

Educational Testing Service. (n.d.). Step III: Manual and technical report. Reading, MA: Addison, Wesley Publishing Company, Inc.

Stanford Achievement Tests: Copperman, P. (1978). The literacy hoax: The decline of reading, writing and learning in the public schools and what we can do about it. New York: William Morrow.

Test Department, Harcourt Brace Jovanovich, Inc. (1983). Some comments on the relationship between scores on the 1973 and 1982 editions of Stanford Achievement Tests. Special Report Number 4A. New York: Harcourt Brace Jovanovich, Inc.

Tests of Achievement and Proficiency:

Riverside Publishing Company. (1983). The development of the 1982 norms for the Iowa Tests of Basic Skills, cognitive abilities test, and tests of achievement and proficiency. Chicago: Riverside Publishing Company.